A New Class of Estimators for the N-point Correlations

István Szapudi[1] and Alexander S. Szalay[2]

[1]NASA/Fermilab Astrophysics Center, Fermi National Accelerator Laboratory, Batavia, IL 60510-0500

[2] Johns Hopkins University, Baltimore, MD 21218

Abstract

A class of improved estimators is proposed for N-point correlation functions of galaxy clustering, and for discrete spatial random processes in general. In the limit of weak clustering, the variance of the unbiased estimator converges to the continuum value much faster than with any alternative, all terms giving rise to a slower convergence exactly cancel. Explicit variance formulae are provided for both Poisson and multinomial point processes using techniques for spatial statistics reported by Ripley (1988). The formalism naturally includes most previously used statistical tools such as N-point correlation functions and their Fourier counterparts, moments of counts-in-cells, and moment correlators. For all these, and perhaps some other statistics our estimator provides a straightforward means for efficient edge corrections.

keywords large scale structure of the universe – galaxies: clustering – methods: numerical – methods: statistical

1. Introduction

As correlation functions are some of the most useful descriptors of galaxy clustering, their accurate estimation is of utmost importance. There is a considerable spread in opinions on what defines an optimal estimator. Since the early work of Peebles (1980 and references therein) and coworkers (e.g. Peebles & Groth 1975, Fry & Peebles 1978) the simple $DD/RR-1$ estimator was widely used, where DD symbolically denotes the number of galaxy pairs at a given range of separations, and RR denotes the number of random pairs generated over a similar area as the data. It has been known for some time that data close to the edges of a survey should have different weights, and for angular correlations an improved estimator has been introduced: $DD/DR-1$, where DR denotes pairs of random and data points. This method computes the contributions from galaxies near the edge in a prorated fashion, via the cross correlations (Hewett 1982). Landy

and Szalay (1993, hereafter LS) have shown that by using $(DD - 2DR + RR)/RR$ one can carry the above argument even further. In the limit of weak clustering the variance is proportional to $1/n^2$, i.e. Poissonian in the pair counts, whereas for all the other estimators the leading order term is $1/n$, where n is the number of points in the survey. Feldman, Kaiser, & Peacock 1994 employed the Fourier equivalent of this expression, and Hamilton 1993 advocated $DD.RR/DR^2$, which, except for a small bias, behaves like the LS estimator. Bernstein 1994 has generalized the LS estimator by including explicitly the effects from higher order clustering. Besides of these direct estimators ensemble estimators were in use as well. Limber 1954, Neyman, Scott, & Shane 1956, Groth & Peebles 1977 used $\langle (N_1 - \langle N \rangle)(N_2 - \langle N \rangle) \rangle \,/\, \langle N \rangle^2$, which is essentially equivalent to the LS estimator. They did not emphasize the difference in the variance of equivalent unbiased estimators, and to date other forms such as $\langle N_1 N_2 \rangle \,/\, \langle N \rangle^2 - 1$, and $\langle N_1 N_2 \rangle \,/\, \langle N_1 \rangle \langle N_2 \rangle - 1$ became wide spread, despite the extra variance they contain. In the statistics literature, probably the most comprehensive review of the pair estimators is by Ripley 1988, who discusses several ways of performing the edge corrections.

At the same time, the estimators for the higher order correlation functions are less well understood. The first theoretical attempt in understanding variance of higher order correlations was Colombi, Bouchet & Schaeffer 1995, where they calculated the variance of the void probability function, P_0. Recently Szapudi & Colombi 1996 identified and calculated the contributions to the variance of moments of counts in cells from the finite survey size, shot noise, the geometry of the survey, as well as measurement effects arising from the finite number of sampling cells. These theoretical computations were applied for realistic survey properties, and the effects of sampling were discussed in Colombi, Szapudi, & Szalay 1997, while the results were extended for the cumulants and for the weakly non-linear regime by Szapudi, Colombi, & Bernardeau 1997. Finally, Matterese, Verde, & Heavens 1997 discussed the variance of the bispectrum. The importance and distinguishing power of the higher order functions depends on how accurately they can be measured, thus any gain in lowering the variance is important.

The main goals of this *Letter* are (i) to explain the advantageous properties of the LS type estimators in simple terms, (ii) generalize them for arbitrary high order, and for any statistics depending on N-tuplets of discrete point processes, (iii) strongly advocate their use instead of traditional estimators with larger variance. We show that for this class of estimators the variance from shot noise is Poissonian in the number of N-tuplets, i.e. $1/n^N$. §2 discusses the physical reasons, why these estimators are superior to earlier ones. In §3 a short but exact derivation of the results is provided for Poissonian and multinomial point processes. The latter represents a survey with a given number of galaxies in it. The importance of the results, possible generalizations, and practical issues are discussed in §4.

2. Minimal Variance Estimators

According to the previous summary of estimators for second order processes the $(DD - 2DR + RR)/RR$ estimator has superior shot noise behavior compared to the existing

alternatives. While the calculations in the next section prove this statement, here a simple argument is given.

The usual $\hat{\xi} = DD/RR - 1$ estimator can be expressed by the sample average of δ, the dimensionless overdensity, as

$$\hat{\xi} = \langle (1+\delta_1)(1+\delta_2) \rangle_s - 1 = \langle \delta_1 + \delta_2 + \delta_1\delta_2 \rangle_s \,, \tag{1}$$

where $\langle \ldots \rangle_s$ denotes the sample average. This is an unbiased estimator, since in the ensemble average $\langle \delta \rangle = 0$, however, the presence of the linear δ terms will add to the variance. Using only the absolutely necessary term, $\delta_1\delta_2$, will obviously decrease the variance. The LS estimator on the other hand can be rewritten as $\hat{\xi}_2 = (D_1 - R_1)(D_2 - R_2)/RR = \langle \delta_1\delta_2 \rangle_s$, i.e. it has the simplest possible structure in δ. This estimator and its Fourier counterpart give the smallest variance, since any other estimator contains extra terms, therefore the total variance can be expressed, as the variance of the LS estimator, plus extra terms,

$$\mathrm{Var}\,\hat{\xi} = \left\langle (\delta_1 + \delta_2)^2 + 2\delta_1^2\delta_2 + 2\delta_1\delta_2^2 \right\rangle_s + \mathrm{Var}(\delta_1\delta_2). \tag{2}$$

Although here we do not attempt to prove, that the extra term is indeed positive, the next section shows mathematically that the LS estimator has minimal shot noise behavior.

Ripley 1988 has discussed extensively the Poisson variance of second order point processes. He has shown, that the term proportional to $1/n$ in the variance of the simple estimator $\hat{\xi}_0$, $\hat{K}_0$ in Ripley's notation, is also proportional to u, the perimeter for a two dimensional survey. This implies that the effect is due to inadequate edge-corrections, in agreement with Hewett's (1982) original suggestion. The subtraction of the appropriate DR terms is equivalent to an optimal edge-correction.

The effect of the unnecessary terms in the estimator on the variance is even more pronounced for the higher order functions, since there will be a lot more terms arising through various combinatorial expressions. Here we propose, that the obvious generalization for higher order correlations is to create the higher order equivalents of the LS estimator, corresponding to $\langle \delta_1 ... \delta_N \rangle$. In the symbolic notation, this estimator can be written as

$$\hat{\xi}_N = (D_1 - R_1).(D_2 - R_2)...(D_N - R_N)/R_1..R_N, \tag{3}$$

with the exact meaning discussed in the next section in rigorous mathematical fashion. This estimator differs from the classic estimator proposed by e.g. Peebles & Groth 1975 for the 3-point function $(DDD - DDR)/RRR + 2$ which contains extra variance for the above explained reasons. Note, however, that the counts-in-cells estimators proposed by Peebles 1975 for the three point function and by Peebles 1980 for the four-point function are the counts-in-cells equivalent of the above equation. It is worth to emphasize again, that for both counts in cells and direct estimators, the above form eliminates the excess variance. On the other hand this form requires some correction to obtain the irreducible, or "connected" correlations. These generalizations will be discussed in the last section.

It will be shown next in a general fashion that the above form constitutes an optimal estimator in the Poisson and multinomial limits. The elegant formalism outlined in Ripley 1988 enabled us to perform most of the calculations in a compact and general form. This work is highly recommended as a reference for mathematical details omitted here.

3. Derivation of the Edge Corrected Estimator

Let D be a catalog of data points to be analyzed, and R randomly generated over the same area, with averages λ, and ρ respectively. The role of R is to perform a Monte Carlo integration compensating for edge effects, therefore eventually the limit $\rho \rightarrow \infty$ will be taken. λ on the other hand is assumed to be externally estimated with arbitrary precision. Many interesting statistics, such as the N-point correlation functions and their Fourier analogs, can be formulated as functions over N-points from the catalog. The covariance of a pair of such estimators will be calculated in the Poisson and multinomial limits. They correspond to the cases, where the number of detected objects is varied or fixed *a priori*. Finally the results are given for the general case, where correlations are non-negligible.

Let us define symbolically an estimator $D^p R^q$, with $p + q = N$ for a function Φ symmetric in its arguments

$$D^q R^p = \sum \Phi(x_1, \ldots, x_p, y_1, \ldots, y_q), \tag{4}$$

with $x_i \neq x_j \in D, y_i \neq y_j \in R$. For example for the two point correlation function $\Phi(x, y) = [x, y \in D, r \leq d(x, y) \leq r + dr]$, where $d(x, y)$ is the distance between the two points, and [*condition*] equals 1 when *condition* holds, 0 otherwise. Ensemble averages can be estimated via factorial moment measures, ν_s (Daley & Vere-Jones 1972, Ripley 1988). In the Poisson limit $\nu_s = \lambda^s \mu_s$, where μ_s is the s dimensional Lebesgue measure.

The general covariance of a pair of estimators is

$$\langle D_a^{p_1} R_a^{q_1} D_b^{p_2} R_b^{q_2} \rangle = \sum_{i,j} \binom{p_1}{i} \binom{p_2}{i} i! \binom{q_1}{j} \binom{q_2}{j} j! \, S_{i+j} \lambda^{p_1+p_2-i} \rho^{q_1+q_2-j}, \tag{5}$$

with

$$S_k = \int \Phi_a(x_1 \ldots x_k, y_{k+1} \ldots y_N) \Phi_b(x_1 \ldots x_k, z_{k+1} \ldots z_N) \mu_{2N-k} \tag{6}$$

Here a and b denote possibly two different radial bins, or even different statistics. The expression describes the i-fold degeneracy in the $p_1 + p_2$ data points from D, as well as the j-fold degeneracy in the $q_1 + q_2$ random points drawn from R. For each of these configurations the geometric phase-space S_{i+j} is different, and the shot noise contribution of each is appropriately summed. The dependence of S_k on a, b, and N is not noted for convenience, but they will be assumed throughout the paper. An estimator for the generalized N-point correlation function is

$$w_N = \frac{1}{S} \sum_i \binom{N}{i} (-)^{N-i} (\frac{D}{\lambda})^i (\frac{R}{\rho})^{N-i}, \tag{7}$$

where $S = \int \Phi\mu_N$ (without subscript). This definition can be expressed as $(\hat{D} - \hat{R})^N$, where $\hat{}$ means normalization with λ^N, ρ^N respectively. In this symbolic Nth power, each factor is evaluated at a different point. Simple calculation in the limit of zero correlations yields $\langle w_N \rangle = 0$. For the same reason, the disconnected parts did not have to be subtracted, which is an important simplification in the calculation for $N \geq 4$. The covariance between two estimators can be evaluated as

$$\langle w_{a,N} w_{b,N} \rangle = \sum_{i_1, i_2, i, j} \binom{N}{i_1} \binom{N}{i_2} \binom{i_1}{i} \binom{i_2}{i} i! \binom{N - i_1}{j} \binom{N - i_2}{j} j! \frac{S_{i+j}}{S^2} \lambda^{2N-i} \rho^{-j} (-)^{2N - i_1 - i_2}. \quad (8)$$

In the interesting limit, where $\rho \to \infty$ only $j = 0$ survives. Changing the order of summation yields

$$\langle w_{a,N} w_{b,N} \rangle = \frac{1}{S^2} \sum_i S_i \lambda^{-i} i! f_{Ni}^2, \quad (9)$$

with

$$f_{Ni} = \sum_j \binom{N}{j} \binom{j}{i} (-1)^{N-j}. \quad (10)$$

This latter can be identified as the coefficients of $\sum_N (xy)^N$, therefore $f_{Ni} = \delta_{Ni}$. Since $\langle w_N \rangle = 0$, the final result is

$$(\mathrm{co})\mathrm{Var}\, w_N = \frac{S_N N!}{S^2 \lambda^N}. \quad (11)$$

Note that this formula represents both variance and covariance depending on whether in the definition of S_N the implicit indices a and b are equal or not.

While in the above Poisson model the total number of galaxies in the survey can vary, it is fixed in the multinomial model. This latter case corresponds to surveys, that detect a certain number of galaxies, and use that to estimate also the mean density. i.e. estimator becomes conditional given the number of galaxies. This introduces some correlations compared to Poisson, which can be taken into account for further precision, especially when the number of galaxies in the survey is relatively small. The normalization of the proper estimator changes slightly: $\lambda^i \to (n)_i / v^i$, where n is the total number of objects in the survey. This normalization renders the estimator unbiased, even when an external estimator for the mean is inavailable. On the other hand, the above Poisson estimator has a slight bias, to leading order $\propto \mathcal{O}(N(N+1)/2n)$, if an internal estimator for the mean is used. For a multinomial process the factorial moment measure is $(n)_N v^{-N} \mu_N$, with v, the volume of the survey, and $(n)_N = n(n-1)\ldots(n-N+1)$, the N-th falling factorial. The covariance

$$\langle w_{a,N} w_{b,N} \rangle = \frac{1}{S^2} \sum_{i, i_1, i_2} S_i v^i i! \binom{N}{i_1} \binom{N}{i_2} \binom{i_1}{i} \binom{i_2}{i} \frac{(n)_{i_1 + i_2 - i} (-)^{i_1 + i_2}}{(n)_{i_1} (n)_{i_2}}. \quad (12)$$

After further reduction (which will be presented elsewhere: Szapudi & Szalay 1997b) the final result is

$$(\mathrm{co})\mathrm{Var}\, w_N = \frac{1}{S^2} \binom{n}{N}^{-1} \sum_i S_i v^i \binom{N}{i} (-)^{N-i}. \quad (13)$$

For $N = 2$ this coincides with LS, taking into account that $S^2 = S_0 \simeq (S_2 v^2)^2$ up to the integral constraint, which was neglected during the previous calculation. Again, the variance is inversely proportional to the number of possible N-tuplets, i.e. Poisson.

4. Discussions

The meaning of the proposed estimator can be understood by specifying the function Φ, such that it is 1 when the N-tuplet satisfies a certain geometry (with a suitable bin width), and 0 otherwise. In this case the estimator will yield the total (or disconnected) N-point correlations of the fluctuations of the underlying field δ. For $N \geq 4$ these contain extra terms compared to the connected correlations, but not as many as the full correlations of the field $1 + \delta$. However, this is the class of estimators which can be precisely corrected for edge effects. The connected correlations can be found by subtracting all the possible partitions. Although this is not simple to take into account in a general fashion it is feasible for any concrete estimator. For example for $N = 4$ the connected part is $\langle \delta_1 \delta_2 \delta_3 \delta_4 \rangle_c = \langle \delta_1 \delta_2 \delta_3 \delta_4 \rangle - (\langle \delta_1 \delta_2 \rangle \langle \delta_3 \delta_4 \rangle + sym.)$. The necessary subtraction will induce extra variance compared to the original estimator, however, its order will still remain λ^{-N}. Partitioning keeps powers of δ unchanged thus it does not introduce lower order terms according to the arguments given in §2; the constant of proportionality might change though. In summary, the connected version of our estimator for the N-point correlations will still have a Poisson variance in the Poisson limit, i.e. it is fully corrected for edge effects.

While so far our discussion focused on the N-point correlation functions, the general nature of the assumed function Φ can incorporate many of the statistics used in the literature, and beyond: Fourier correlation functions ($N - 1$ spectra), moments of counts in cells, and factorial moment correlators are obvious candidates. For the sake of illustration, let us briefly mention the suitable Φ functions, which correspond to the above quantities. For the $N - 1$ spectra $\Phi = \sum_{\sigma(x_1,\ldots,x_N)} \Pi e^{i x_j k_j}/\mathrm{N}!$, where σ refers to the sum over all possible permutations of the points x_i to render the function symmetric. Possibly, weights can be incorporated as well. Then the definition for $N = 2$ is equivalent to Feldman, Kaiser, & Peacock 1994 up to the shot noise power, which is not present in our estimator, since overlapping indices are conveniently excluded in the definition. Moments of counts-in-cells can be represented similarly. The N-th factorial moment in a cell is estimated by $\Phi = 1$ if a certain N-tuplet *fit* into the cell, 0 otherwise. For $N = 2$ this method will estimate Ripley's K_0 function, or the average of the two-point function over a cell. Our estimator therefore will give $\left\langle \delta^N \right\rangle$, from which the cumulants $Q_N \propto \left\langle \delta^N \right\rangle_c$ can be straightforwardly calculated. Finally, if $\Phi = 1$ for a set of $N + M$ points fit into a pair of cells separated by distance r, and 0 otherwise, factorial moment correlators are estimated. The proposed estimator gives $\left\langle \delta_1^N \delta_2^N \right\rangle$, from which the cumulant correlators (Szapudi & Szalay 1997a) are simply obtained as $Q_{NM} \propto \left\langle \delta_1^N \delta_2^N \right\rangle_c$. Generalization for k-th order joint moments is obvious.

The fact that the estimator is not exactly connected is of little importance for the N-point correlations and their Fourier counterparts. It is unlikely that in the near future these methods can be pushed much beyond $N = 4$, and the calculation of the N-point function from the

estimator of Equation 3 is quite straightforward. The other two statistics can be measured up to higher order: it was demonstrated both in simulations (Colombi, Bouchet, & Hernquist 1995, Baugh, Gaztañaga, & Efstathiou 1995) and galaxy catalogs (Peebles 1980, Szapudi, Szalay & Boschán 1992, Gaztañaga 1992, Bouchet *et al.* 1993, Meiksin, Szapudi, & Szalay 1992, Gaztañaga 1994, Szapudi *et al.* 1995, Szapudi, Meiksin, & Nichol 1996), that the moments can be extracted up to 10th order. However, the connected moments, that is cumulants and cumulant correlators (Szapudi & Szalay 1997a), can be calculated simply via generating functions from our estimator. Since the correction for connected moments does not introduce lower order terms in the variance, the resulting statistics will be better corrected for edge effects, than any previous methods based on moments of counts in cells, which cannot be corrected for edge effects (Szapudi & Colombi 1996, Colombi, Szapudi, & Szalay 1997).

So far we proved rigorously that the proposed edge correction is valid for a Poisson distribution. The calculation shows how most of the shot noise is eliminated during this process. The eliminated discreteness error is also an edge effect, and it can be corrected for. We propose the term "edge-discreteness" effect, to further refine the classification of Szapudi & Colombi 1996. Since previous estimators were not corrected for this extra shot noise-variance, we conjecture that they fare worse even when correlations are present. Although we did not prove this rigorously (that would require the generalization of Bernstein 1994 for higher order correlations), the arguments of §2 show that this is the case. The actual formulae for the errors, however, can only be applied to estimate the discreteness contribution to the error (usually not dominant at large scales except for small fields with deep exposures, as HDF, etc.) since it does not take into account the finite volume and edge effects. These are always present in a realistic distribution and depend on the integral of the correlation function, and on the two-point and higher order correlation functions (Szapudi & Colombi 1996). Next we outline possible generalizations, which could improve on the approximations if necessary by taking into account the correlations when performing the ensemble averages.

As explained above, the explicit variance formulae can be applied reliably when the finite volume and edge effects are not dominating (Szapudi & Colombi 1996), i.e. a sparsely sampled survey, where the integral of the correlation function over the survey volume is small as well. Fortunately, edge effects are mostly eliminated by the proposed estimator, while finite volume effects can be estimated by actually evaluating the correlation integral over the survey volume. However, a more accurate calculation is possible, using factorial moment measures which include all the needed higher order correlations, or possibly truncated at the appropriate order if large scales are considered. For the variance of the generalized N-point correlation function, a model for up to $2N$ statistics is needed. Such a calculation for the 2-point function was performed by Bernstein 1994. Note however, that any estimation of the variance on higher order correlation functions is approximate, because all the systematics, and inaccuracy of the prior model will be geometrically amplified, in addition to the fact that the interpretation of the variance becomes less and less straightforward as the error distribution deviates from the Gaussian (Szapudi & Colombi 1996). Nevertheless the calculation is possible although tedious and the basic steps are outlined below.

For a general point process the factorial moment measures can be expressed as $\nu_N = F_N \mu_N$, where F_N's are the reducible N-point correlation functions. Once this is given, it is only a matter of simple but lengthy calculations to express all the ensemble averages needed, using the fact that the scaling of the factorial moment measure with λ for a general (canonical) point process is identical to the Poisson case. The result is quite similar to Equation 12, except the falling factorials and the powers of the volume are replaced simply by λ^{-i}, and S_i will depend additionally on the indices i_1 and i_2. Some simplification can be achieved by expressing $F_N = \lambda^N(1 + \xi + \ldots + \xi_3 + \ldots + \xi_N)$ with the irreducible correlation functions, and, because of the assumed symmetry of the functions Φ, replacing by $\lambda^N \sum_i C_i^N \xi_i$. The C_i^N's are symmetry factors calculable from generating functions (Szapudi & Szalay 1993) or cluster expansion, and $C_1 = C_N = 1, \xi_1 = 1$. This makes it possible to replace the integrals $S_i^{i_1 i_2}$ with irreducible integrals. It would be beyond the scope of this work to quote the explicit formula and explore its applicability. Our goal was only to clearly show how such calculation can be trivially performed if necessary, the rest is left for subsequent research.

This *Letter* proposed a set of edge corrected estimators for the generalized N-point correlation functions. The estimators were defined using the framework provided by recent results in spatial statistics, which can handle in a uniform way almost all previously used statistics for characterizing higher order clustering. Thus our formalism includes among others the N-point correlation functions, N-point Fourier transforms (or $N-1$-spectra), moments of counts-in-cells, and moment correlators. Explicit calculations were performed in the Poisson and multinomial limits to show that the variance is approximately Poisson, i.e. in both cases it is inversely proportional to the number of possible N-tuplets. The calculation also pinpointed how previous estimators were not even corrected for the "edge-discreteness" effects, and physical arguments were given, that they would perform even worse when correlations are present. We proposed ways to estimate the other important contributions, the edge and finite volume effects, thus a combination of numerical estimates combined with our formulae will yield substantially improved accuracy over previous techniques.

I.S. would like to thank L. Hui and S. Dodelson for discussions. I.S. was supported by DOE and NASA through grant NAG-5-2788 at Fermilab. A.S.Z. was supported by a NASA LTSA grant.

REFERENCES

Baugh, C.M., Gaztañaga, E., Efstathiou, G. 1995, MNRAS, 274, 1049

Bernstein, G. M. 1994, ApJ, 434, 569

Bouchet, F.R., Strauss, M.A., Davis, M., Fisher, K.B., Yahil, A., & Huchra, J.P. 1993, ApJ, 417, 36

Colombi, S., Bouchet, F.R., & Hernquist, L. 1995, A&A, 281, 301

Colombi, S., Bouchet, F.R., & Schaeffer, R. 1995, ApJS, 96, 401

Colombi, S., Szapudi, I. Szalay, A.S. 1997, in preparation

Daley, D.J., & Vere-Jones, D. 1972 In *Stochastic Point Processes*, Ed. Lewis, P.A.W. (New York: Wiley) in Rippley

Feldman, H. A., Kaiser, N., & Peacock, J. A. 1994, ApJ, 426, 23

Fry, J.N., & Peebles, P.J.E. 1978, ApJ, 221, 19

Gaztañaga, E. 1992, ApJ, 319, L17

Gaztañaga, E. 1994, MNRAS, 268, 913

Groth, E.J., & Peebles, P.J.E. 1977, ApJ, 217, 385

Hamilton, A.J.S. 1993, ApJ, 417, 19

Hewett, H.C. 1982, MNRAS, 201, 867

Landy, S.D.,& Szalay, A. 1993, ApJ, 412, 64

Limber, D.N., 1954, ApJ, 119, 655

Matterese, S., Verde, L., & Heavens, A.F. 1997, in preparation

Meiksin, A., Szapudi, I., & Szalay, A., 1992, ApJ, 394, 87

Neyman, J., Scott, E.L.,& Shane, C.D 1956 Proc Third Berkeley Simposium Math Stat and Probability. 3, 75

Peebles, P.J.E., 1975, ApJ, 196, 647

Peebles, P.J.E. 1980, The Large Scale Structure of the Universe (Princeton: Princeton University Press)

Peebles, P.J.E., & Groth, E.J. 1975, ApJ, 196, 1

Ripley, B.D. 1988, Statistical Inference for Spatial Processes (Cambridge: Cambridge Univ. Press)

Szapudi, I., & Colombi, S. 1996, ApJ, 470, 131

Szapudi, I., Colombi, S., Bernardeau, F. 1997, in preparation

Szapudi, I., Meiksin, A., & Nichol, R.C. 1996, ApJ, 473, 15

Szapudi, I. & Szalay, A. 1993, ApJ, 408, 43

Szapudi, I., Szalay, A.S. 1997a, accepted, ApJ

Szapudi, I., Szalay, A.S. 1997b, in preparation

Szapudi, I., Dalton, G., Efstathiou, G.P., & Szalay, A. 1995, ApJ, 444, 520

Szapudi, I., Szalay, A., & Boschán, P. 1992, ApJ, 390, 350

This preprint was prepared with the AAS LaTeX macros v4.0.